IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Ludmila Cherkasova, et al.	§	Art Unit:	2153
		§		
Serial No.:	10/660,978	§		
		§	Examiner:	T. K. Kim
Filed:	September 12, 2003	§		
		§		
For:	System and Method for	§	Atty. Dkt. No.:	200313317-1
	Evaluating a Capacity of a	§		(HPC.0555US)
	Streaming Media Server for	§		
	Supporting a Workload	§		

Mail Stop Appeal Brief-Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

APPEAL BRIEF PURSUANT TO 37 C.F.R § 41.37

Sir:

The final rejection of claims 1-35, 37-39, and 42-44 is hereby appealed.

I. REAL PARTY IN INTEREST

The real party in interest is the Hewlett-Packard Development Company, LP.

II. RELATED APPEALS AND INTERFERENCES

None.

III.   STATUS OF THE CLAIMS

Claims 1-35, 37-39, and 42-44 have been finally rejected and are the subject of this appeal.

Claims 36, 40, and 41 have been cancelled.

IV.   STATUS OF AMENDMENTS

An Amendment Under 37 C.F.R. § 41.33 is submitted concurrently herewith to cancel claims 40 and 41. Entry of the Amendment is proper since the Amendment merely cancels claims without affecting the scope of other pending claims. *See* § 41.33(b).

V.   SUMMARY OF THE CLAIMED SUBJECT MATTER

The following provides a concise explanation of the subject matter defined in each of the independent claims involved in the appeal, referring to the specification by page and line number and to the drawings by reference characters, as required by 37 C.F.R. § 41.37(c)(1)(v). Each element of the claims is identified by a corresponding reference to the specification and drawings where applicable. Note that the citation to passages in the specification and drawings for each claim element does not imply that the limitations from the specification and drawings should be read into the corresponding claim element.

Independent claim 1 recites a method comprising:

receiving (Fig. 9:901), into a capacity planning tool (Fig. 1:101), configuration information (Fig. 1:103) for at least one streaming media server, wherein the configuration information comprises a single file benchmark and a unique file benchmark for the at least one streaming media server (Spec., pp. 10-11, ¶ [0038]; p. 20, ¶ [0065]).

receiving (Fig. 9:902), into said capacity planning tool, workload information for a workload of client accesses of streaming media files from a server (Spec., p. 20, ¶ [0065]); and

said capacity planning tool evaluating (Fig. 9:903), based on said configuration information, a capacity of the at least one streaming media server for supporting the workload (Spec., pp. 12-13, ¶ [0043]; p. 20, ¶ [0065]).

Independent claim 10 recites a method comprising:

receiving (Fig. 9:901), into a capacity planning tool (Fig. 1:101), configuration information for at least one streaming media server (Spec., p. 20, ¶ [0065]);

receiving (Fig. 9:902), into said capacity planning tool, workload information for a workload of client accesses of streaming media files from a server (Spec., p. 20, ¶ [0065]);

said capacity planning tool evaluating (Fig. 9:903) a capacity of the at least one streaming media server for supporting the workload (Spec., pp. 12-13, ¶ [0043]; p. 20, ¶ [0065]);

wherein said evaluating comprises computing a service demand for said at least one streaming media server supporting said workload (Spec., p. 13, ¶ [0044]); and

wherein said computing said service demand comprises computing (Spec., p. 13, ¶ [0045]):

$$Demand = \sum_{i=1}^{K_W} N_{X_{W_i}}^{memory} \times cost_{X_{W_i}}^{memory} + \sum_{i=1}^{K_W} N_{X_{W_i}}^{disk} \times cost_{X_{W_i}}^{disk} \, ,$$

wherein the workload W comprises $X_W = X_1,..., X_{k_w}$ set of different encoded bit rates of files served in the workload, $N_{X_{W_i}}^{memory}$ is a number of streams in the workload having a memory access to a subset of files encoded at X_{W_i} Kb/s, $cost_{X_{W_i}}^{memory}$ is a cost of consumed resources for a stream having a memory access to a file encoded at X_{W_i} Kb/s, $N_{X_{W_i}}^{disk}$ is a number of streams in the workload having a disk access to a subset of files encoded at X_{W_i} Kb/s, and $cost_{X_{W_i}}^{disk}$ is a cost of consumed resources for a stream having a disk access to a file encoded at X_{W_i} Kb/s (Spec., p. 13, ¶ [0044]).

Independent claim 16 recites computer-executable software code (Fig. 1:101) stored to a computer-readable medium, the computer-executable software code comprising:

code for receiving workload information for a workload of client accesses of streaming media files from a server (Spec., p. 20, ¶ [0065]); and

code for employing a cost function derived for at least one system configuration from a single file benchmark and a unique file benchmark for evaluating a capacity of the at least one system configuration for supporting the workload (Spec., pp. 10-11, ¶ [0038]; pp. 12-13, ¶ [0043]; p. 20, ¶ [0064]).

Independent claim 25 recites a system comprising:

means for receiving configuration information for a plurality of different system configurations, wherein the configuration information comprises, for each of the plurality of different system configurations, a corresponding single file benchmark and unique file benchmark, wherein said single file benchmark measures capacity of the corresponding system configuration for serving a population of clients that all access a same file, wherein said unique file benchmark measures capacity of the corresponding system configuration for serving a population of clients that all access different files (Spec., pp. 10-11, ¶ [0038]; p. 20, ¶ [0065]);

means for receiving workload information for a workload of client accesses of streaming media files from a server (Spec. p. 20, ¶ [0065]); and

means for evaluating, based on the configuration information, the capacity of each of said plurality of different system configurations for supporting said workload (Spec. p. 20, ¶ [0065]).

Independent claim 28 recites a method comprising:

receiving (Fig. 9:902) workload information identifying an expected workload of client accesses of streaming media files from a server over a period of time (Spec., pp. 4-5, ¶ [0024]; p. 20, ¶ [0065]); and

determining a service demand profile for at least one server configuration under evaluation for evaluating a capacity of said at least one server configuration for supporting the expected workload, wherein said service demand profile comprises a plurality of pairs of information, each pair comprising an identification of a duration of time in said period of time and a corresponding computed resource cost of the at least one server configuration for serving the workload over the duration of time (Spec., pp. 17-18, ¶ [0059]).

Independent claim 32 recites a system comprising:

a media profiler (Fig. 2:202) operable to receive a client access log collected over a period of time for a service provider's site and generate a workload profile for the service provider's site, wherein said workload profile comprises, for a plurality of different points in time, identification of a number of concurrent client accesses, wherein the number of concurrent client accesses are categorized into corresponding encoding bit rates of streaming media files accessed thereby and are further sub-categorized into either

4

memory or disk accesses (Spec., p. 15, ¶ [0052]); and

a capacity evaluator (Fig. 1, 101) operable to receive the generated workload profile and evaluate at least one server configuration's capacity for supporting the site's workload (Spec., p. 20, ¶ [0064]).

Independent claim 44 recites a method comprising:

determining results of a single file benchmark for each of a plurality of encoding bit rates of a single file served by at least a first streaming media server configuration, wherein the result of the single file benchmark for a given encoding bit rate identifies the maximum number of concurrent streams of the single file that the at least a first streaming media server configuration can supply to a population of clients at the given encoding bit rate (Spec., p. 11, ¶ [0039]);

determining results of a unique file benchmark for each of said plurality of encoding bit rates, wherein the result of the unique file benchmark for a given encoding bit rate identifies the maximum number of concurrent streams of different files that the at least a first streaming media server configuration can supply to the population of clients at the given encoding bit rate (Spec., p. 11 ¶ [0040]);

deriving, from the results of the single file benchmark and unique file benchmark, a cost function (Spec., p. 12, ¶ [0043]);

receiving, into a capacity planning tool, workload information for a workload of client accesses of streaming media files from a server (Spec., p. 19, ¶ [0063]); and

using, by the capacity planning tool, the cost function for said at least a first streaming media server configuration for evaluating a capacity of the at least a first streaming media server configuration for supporting the workload (Spec., p. 20, ¶ [0064]).

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL[1]

A. **Claims 28-35, 37, 38, and 43 rejected under 35 U.S.C. § 102(e) as being anticipated by Lumelsky (U.S. Patent No. 6,516,350).[2]**

B. **Claims 1-9, 11-27, and 42 rejected under 35 U.S.C. § 103(a) as being unpatentable over Lumelsky in view of Haroldson (U.S. Application No. 2002/0029373).**

C. **Claims 10, 39, and 44 rejected under 35 U.S.C. § 103(a) as being unpatentable over Lumelsky, in view of Haroldson, and in further view of Drees (U.S. Patent No. 5,778,683).**

VII. ARGUMENT

The claims do not stand or fall together. Instead, Appellant presents separate arguments for various independent and dependent claims. Each of these arguments is separately argued below and presented with separate headings and sub-headings as required by 37 C.F.R. § 41.37(c)(1)(vii).

A. Claims 28-35, 37, 38, and 43 rejected under 35 U.S.C. § 102(e) as being anticipated by Lumelsky (U.S. Patent No. 6,516,350).

1. Claims 28-31.

It is respectfully submitted that claim 28 is clearly not anticipated by Lumelsky.

Claim 28 recites a method comprising:

- receiving workload information identifying an expected workload of client accesses of streaming media files from a server over a period of time; and

- determining a service demand profile for at least one server configuration under evaluation for evaluating a capacity of said at least one server configuration for supporting the expected workload, wherein said service demand profile comprises a plurality of pairs of information, each pair comprising an identification of a duration of time in said period of time and a corresponding computed resource

---

cost of the at least one server configuration for serving the workload over the duration of time.

Lumelsky clearly does not disclose a service demand profile that comprises a plurality of pairs of information, where each pair comprises an identification of a duration of time in the period of time and a corresponding computed resource cost of the at least one server configuration for serving the workload over the duration of time.

With respect to this element of claim 28, the Examiner cited Fig. 9 and column 12, lines 2-7 of Lumelsky. 1/25/2008 Office Action at 7. Fig. 9 is a "graphical depiction of the capacity shaping mechanism" described in Lumelsky. Lumelsky, 7:49-51. As noted in column 12 of Lumelsky, a line 661 in Fig. 9 of Lumelsky represents the number of request streams measured per pre-established time interval. The number of request streams measured per pre-established time interval does not constitute the pair of information that comprises an identification of a duration of time in the period of time and a **corresponding computed resource cost of the at least one server configuration for serving the workload over the duration of time.**

Column 12 of Lumelsky goes on to state that a dotted line 662 in Fig. 9 symbolizes the number of possible streams per "this interval." *Id.*, 12:4-5. The number of possible streams per time interval is not the same as **a corresponding computed resource cost of the at least one server configuration for serving the workload over the duration of time**, as recited in claim 28.

There is no other passage in Lumelsky that discloses the subject matter of claim 28. Therefore, claim 28 and its dependent claims are clearly allowable over Lumelsky.

Reversal of the final rejection of the above claims is respectfully requested.

2.   **Claims 32-25, 37.**

Independent claim 32 recites a system comprising:

- a media profiler operable to receive a client access log collected over a period of time for a service provider's site and generate a workload profile for the service provider's site, wherein said workload profile comprises, for a plurality of different points in time, identification of a number of concurrent client accesses, wherein the number of concurrent client accesses are categorized into corresponding encoding bit rates of streaming media files accessed thereby and are further sub-categorized into either memory or disk accesses; and

- a capacity evaluator operable to receive the generated workload profile and evaluate at least one server configuration's capacity for supporting the site's workload.

It is clear that Lumelsky fails to disclose at least the following combination of elements: a workload profile that comprises, for a plurality of different points in time, identification of a number of concurrent client accesses, where the number of concurrent client accesses are **categorized into corresponding encoding bit rates of streaming media files accessed thereby and are further sub-categorized into either memory or disk accesses.** As purportedly disclosing this feature of claim 32, the Examiner cited Figs. 10 and 11 and column 12, lines 36-41, of Lumelsky. 1/25/2008 Office Action at 8. The cited column 12 passage of Lumelsky states that Fig. 10 illustrates streaming of MPEQ-1 content and of MPEG-2 content. The cited passage also notes that "as a convenient means of measurement of the consumption of resources by a [*sic*] various classes of applications, the resources are characterized as storage s and service r(m, c, d) bins." However, there is no indication whatsoever that a number of concurrent client accesses are categorized into corresponding encoding bit rates of streaming media files accessed thereby and are further sub-categorized into either memory or disk accesses.

The Examiner had cited Fig. 9 of Lumelsky as disclosing the identification of a number of concurrent client accesses for a plurality of different points in time. Fig. 9 shows a number of

streams per interval (661) or number of possible streams per interval (662). However, there is absolutely no indication whatsoever that the number of streams per interval or number of possible streams per interval depicted in Fig. 9 are categorized and further sub-categorized in the manner recited in claim 32.

Therefore, claim 32 and its dependent claims are clearly not anticipated by Lumelsky.

Reversal of the final rejection of the above claims is respectfully requested.

3. Claim 43.

Claim 43 depends indirectly from claim 32 and is therefore allowable for at least the same reasons as claim 32. Moreover, claim 43 recites that the service demand profile comprises a plurality of pairs of information, each pair comprising identification of a duration of time in the period of time and a corresponding computed resource cost of the at least one server configuration for serving the workload over the duration of time. For additional reasons similar to those stated above with respect to claim 28, claim 43 is further allowable over Lumelsky.

Reversal of the final rejection of the above claim is respectfully requested.

B. Claims 1-9, 11-27, and 42 rejected under 35 U.S.C. § 103(a) as being unpatentable over Lumelsky in view of Haroldson (U.S. Application No. 2002/0029373).

1. Claims 1-6, 9, 11-15, 38.

It is respectfully submitted that the obviousness rejection of claim 1 over Lumelsky and Haroldson is erroneous. As conceded by the Examiner, Lumelsky fails to disclose configuration information for at least one streaming media server, where the configuration information comprises a single file benchmark and a unique file benchmark for the at least one streaming media server. 1/25/2008 Office Action at 11. Instead, the Examiner cited Haroldson as

purportedly disclosing a single file benchmark and a unique file benchmark for at least one streaming media server, citing specifically to ¶ [0005] of Haroldson.

The cited passage of Haroldson refers to a technique for calculating a number of concurrent network connections associated with multicast streaming media. The cited passage also states that a chart can be generated that illustrates the number of network connections at specific points of time. Moreover, the passage refers to calculating concurrent connections for a particular server and/or a specific data stream. However, nowhere in this passage of Haroldson, or anywhere else in Haroldson, is there any hint given of configuration information that comprises a single file benchmark and a unique file benchmark for the at least one streaming media server. Calculating the number of network connections for a particular server and/or a specific data stream does not constitute a single **file** benchmark and a unique **file** benchmark, as recited in claim 1.

As explained in the specification, according to some embodiments, a single file benchmark measures a media server capacity when all clients in a test workload are accessing the same file, while a unique file benchmark measures a media server capacity when each client in the test workload is accessing a different file. *See, e.g.,* Specification, page 10, ¶ [0038]. Merely calculating the number of concurrent network connections, as taught by Haroldson, does not constitute providing a single file benchmark and a unique file benchmark for at least one streaming media server.

Therefore, even if Lumelsky and Haroldson could be hypothetically combined, the hypothetical combination of the references fails to disclose or hint at all elements of claim 1.

Moreover, contrary to the assertion by the Examiner, Lumelsky fails to disclose or hint at the capacity planning tool evaluating, based on the configuration information, a capacity of the at

least one streaming media server for supporting the workload, as recited in claim 1. Note that claim 1 further recites that the capacity planning tool evaluates, based on the configuration information that comprises a single file benchmark and a unique file benchmark for the at least one streaming media server, a capacity of the at least one streaming media server for supporting the workload. This subject matter of claim 1 is also not disclosed or hinted at by the hypothetical combination of Lumelsky and Haroldson.

Therefore, the obviousness rejection of claim 1 and its dependent claims is erroneous.

Reversal of the final rejection of the above claims is respectfully requested.

2.   Claims 7, 8.

Dependent claim 7 depends from claim 1 and is therefore allowable for at least the same reasons as claim 1. Moreover, claim 7 recites that evaluating a capacity of the at least one streaming media server for supporting the workload based on the configuration information comprises computing a cost corresponding to resources of the at least one streaming media server that are consumed in supporting the workload. The Examiner cited column 10, lines 45-47 and 51-53 of Lumelsky as disclosing this feature of claim 7. The cited passages of Lumelsky refer to a billing and pricing service that enables media services to be offered at various costs and permits a wide variety of pricing policies. The billing and pricing service described in Lumelsky has nothing to do with computing a cost corresponding to resources of the at least one streaming media server in the context of evaluating the capacity of the at least one streaming media server for supporting the workload.

Therefore, claim 7 and its dependent claim are further allowable for the foregoing reasons.

Reversal of the final rejection of the above claims is respectfully requested.

3.   Claims 16-21, 23, 24.

Independent claim 16 recites computer-executable software code that includes code for receiving workload information for a workload of client accesses of streaming media files from a server, and code for employing a cost function derived for at least one system configuration from a **single file benchmark and a unique file benchmark** for evaluating a capacity of the at least one system configuration for supporting the workload.

As explained above, contrary to the assertion of the Examiner, Haroldson fails to disclose or hint at the single file benchmark and the unique file benchmark recited in claim 16, nor does Haroldson provide any teaching of evaluating a capacity of at least one system configuration for supporting a workload by using a cost function derived from a system configuration from the single file benchmark and the unique file benchmark.

Therefore, even if Lumelsky and Haroldson could be hypothetically combined, the hypothetical combination of references would not have led to the subject matter of claim 16.

Therefore, the obviousness rejection of claim 16 and its dependent claims is defective. Reversal of the final rejection of the above claims is respectfully requested.

4.   Claim 22.

Claim 22 depends indirectly from claim 16 and is therefore allowable for at least the same reasons as claim 16. Claim 22 further recites that the workload profile generated for the received workload information comprises, for a plurality of different points in time, identification of a number of concurrent client accesses, where the number of concurrent client accesses are **categorized into corresponding encoding bit rates** of streaming media files accessed thereby and are further **sub-categorized into either memory or disk accesses**. As explained above in connection with claim 32, this is a feature that is clearly not disclosed or

hinted at by Lumelsky, contrary to the assertion by the Examiner. Therefore, even if Lumelsky and Haroldson could be hypothetically combined, the hypothetical combination of the references would not have led to the subject matter of claim 22.

Reversal of the final rejection of the above claim is therefore respectfully requested.

5.   Claims 25-27.

Independent claim 25 is also non-obvious over Lumelsky and Haroldson. Claim 25 recites means for receiving configuration information for a plurality of different system configurations, where the configuration information comprises, for each of the plurality of different system configurations, a corresponding single file benchmark and unique file benchmark, where the single file benchmark measures capacity of the corresponding system configuration for serving a population of clients that all access a same file, and where the unique filed benchmark measures capacity of the corresponding system configuration for serving a population of clients that all access different files.

As discussed above in connection with claim 1, the hypothetical combination of Lumelsky and Haroldson clearly does not teach or hint at the single file benchmark and unique file benchmark of claim 25.

Since the hypothetical combination of the references fails to disclose or hint at the configuration information that includes the unique file benchmark and single file benchmark recited in claim 25, it is also respectfully submitted that the hypothetical combination does not teach or hint at means for evaluating, based on **the** configuration information, the capacity of each of the plurality of different system configurations for supporting the workload.

Therefore, the obviousness rejection of claim 25 and its dependent claims is erroneous. Reversal of the final rejection of the above claims is respectfully requested.

6.   Claim 42.

In view of the defective rejection of base claim 28 over Lumelsky, it is respectfully submitted that the obviousness rejection of claim 42 over Lumelsky and Haroldson is also defective. Moreover, as discussed above in connection with claim 1, Haroldson does not disclose or hint at a single file benchmark and unique file benchmark, as recited in claim 42. Therefore, the hypothetical combination of Lumelsky and Haroldson would not have led to the subject matter of claim 42. The obviousness rejection of claim 42 is therefore defective, and reversal of the final rejection of the above claim is respectfully requested.

C.   Claims 10, 39, and 44 are rejected under 35 U.S.C. § 103(a) as being unpatentable over Lumelsky, in view of Haroldson, and in further view of Drees (U.S. Patent No. 5,778,683).

1.   Claim 10.

Independent claim 10 was rejected as being obvious over Lumelsky, Haroldson, and Drees. As conceded by the Examiner, Lumelsky and Haroldson fail to disclose the equation for calculating *Demand* as recited in claim 10, where the equation for *Demand* is computed based on a number of streams in the workload having a memory access to a subset of files encoded at a corresponding encoded bit rate, a cost of consumed resources for a stream having a memory access to a file encoded at a corresponding encoded bit rate, a number of streams in the workload having a disk access to a subset of files encoded at a corresponding encoded bit rate, and a cost of consumed resources for a stream having a disk access to a file encoded at a corresponding encoded bit rate.

The Examiner erroneously asserted that Drees discloses such an equation for computing the service demand (*Demand*) recited in claim 10. Drees describes commercial customers being typically billed based upon energy consumed and the peak demand. Drees 1:28-30. As

explained by Drees, demand is the average instantaneous power consumed over a one to fifteen minute interval. *Id.*, 1:32-34. The peak demand is the maximum demand incurred over a particular billing period. *Id.*, 1:34-36. Demand costs are computed by multiplying the maximum demand incurred during the billing period by the demand charge. *Id.*, 1:39-41. Drees also notes that demand charges are adjusted based on the time of use to further motivate the customer to shift electrical consumption from higher rate on-peak time periods to lower rate off-peak time periods. *Id.*, 1:41-45. This teaching of Drees has absolutely nothing to do with computing the service demand based on the following factors: the number of streams in the workload having a memory access to a subset of files encoded at a corresponding encoded bit rate, a cost of consumed resources for a stream having a memory access to a file encoded at a corresponding encoded bit rate, a number of streams in the workload having a disk access to a subset of files encoded at a corresponding encoded bit rate, and a cost of consumed resources for a stream having a disk access to a file encoded at a corresponding bit rate.

Therefore, even if Drees could be properly combined with Lumelsky and Haroldson, the hypothetical combination of these references would not have disclosed or hinted at the calculation of the service demand in accordance with claim 10.

Thus, the obviousness rejection of claim 10 is clearly erroneous. Reversal of the final rejection of the above claim is respectfully requested.

2.   Claim 39.

Claim 39 is allowable over the asserted combination of Lumelsky, Haroldson, and Drees for similar reasons as claim 10.

Reversal of the final rejection of the above claim is respectfully requested.

3. Claim 44.

Independent claim 44 was rejected over Lumelsky, Haroldson, and Drees for reasons "stated above per rejection to Claim 40." 1/25/2008 Office Action at 21. However, in the rejection of claim 44, the Examiner failed to cite to the actual language of claim 44 in rejecting claim 44 over the references. Specifically, the Examiner failed to address the "single file benchmark" and "unique file benchmark" elements of claim 44. Therefore, the rejection of claim 44 over the references is defective.

The hypothetical combination of Lumelsky, Haroldson, and Drees clearly fails to disclose or hint at:

- determining results of a **single file benchmark** for each of a plurality of **encoding bit rates** of a single file served by at least a first streaming media server configuration, wherein the result of the **single file benchmark** for a given **encoding bit rate** identifies the maximum number of concurrent streams of the single file that the at least a first streaming media server configuration can supply to a population of clients at the given **encoding bit rate**;

- determining results of a **unique file benchmark** for each of said plurality of **encoding bit rates**, wherein the result of the **unique file benchmark** for a given **encoding bit rate** identifies the maximum number of concurrent streams of different files that the at least a first streaming media server configuration can supply to the population of clients at the given **encoding bit rate**

In view of the foregoing, it is clear that the obviousness rejection of claim 44 is erroneous, and reversal of the final rejection of the above claim is respectfully requested.

CONCLUSION

In view of the foregoing, reversal of all final rejections and allowance of all pending claims is respectfully requested.

Respectfully submitted,

Date: 6/25/2008

Dan C. Hu
Registration No. 40,025
TROP, PRUNER & HU, P.C.
1616 South Voss Road, Suite 750
Houston, TX 77057-2631
Telephone: (713) 468-8880
Facsimile: (713) 468-8883

VIII. APPENDIX OF APPEALED CLAIMS

The claims on appeal are:

1   1.   A method comprising:

2   receiving, into a capacity planning tool, configuration information for at least one

3 streaming media server, wherein the configuration information comprises a single file

4 benchmark and a unique file benchmark for the at least one streaming media server;

5   receiving, into said capacity planning tool, workload information for a workload

6 of client accesses of streaming media files from a server; and

7   said capacity planning tool evaluating, based on said configuration information, a

8 capacity of the at least one streaming media server for supporting the workload.

1   2.   The method of claim 1 wherein said configuration information includes

2 identification of size of memory of said at least one streaming media server.

1   3.   The method of claim 2 wherein said configuration information further includes

2 disk configuration of said at least one streaming media server.

1   4.   The method of claim 1 wherein said workload information includes identification

2 of number of concurrent client accesses of said streaming media files over a period of time.

1   5.   The method of claim 4 wherein said workload information further includes

2 identification of a corresponding encoding bit rate of each of said streaming media files

3 accessed.

1   6.   The method of claim 1 wherein said workload information comprises information

2 from an access log collected over a period of time.

1   7.   The method of claim 1 wherein said evaluating comprises:

2   computing a cost corresponding to resources of said at least one streaming media

3 server that are consumed in supporting the workload.

1   8.   The method of claim 7 wherein said computing said cost comprises:

2   computing a cost of consumed resources for a stream in said workload having a

3   memory access to a streaming media file; and

4   computing a cost of consumed resources for a stream in said workload having a

5   disk access to a streaming media file.

1   9.   The method of claim 1 wherein said evaluating comprises:

2   computing a service demand for said at least one streaming media server

3   supporting said workload.

1   10.   A method comprising:

2   receiving, into a capacity planning tool, configuration information for at least one

3   streaming media server;

4   receiving, into said capacity planning tool, workload information for a workload

5   of client accesses of streaming media files from a server;

6   said capacity planning tool evaluating a capacity of the at least one streaming

7   media server for supporting the workload;

8   wherein said evaluating comprises computing a service demand for said at least

9   one streaming media server supporting said workload; and

10   wherein said computing said service demand comprises computing:

11
$$Demand = \sum_{i=1}^{K_W} N_{X_{W_i}}^{memory} \times \cos t_{X_{W_i}}^{memory} + \sum_{i=1}^{K_W} N_{X_{W_i}}^{disk} \times \cos t_{X_{W_i}}^{disk} \; ,$$

12   wherein the workload W comprises $X_W = X_1, ..., X_{k_w}$ set of different encoded bit

13   rates of files served in the workload, $N_{X_{W_i}}^{memory}$ is a number of streams in the workload having a

14   memory access to a subset of files encoded at X_{W_i} Kb/s, $\cos t_{X_{W_i}}^{memory}$ is a cost of consumed

15   resources for a stream having a memory access to a file encoded at X_{W_i} Kb/s, $N_{X_{W_i}}^{disk}$ is a number

16   of streams in the workload having a disk access to a subset of files encoded at X_{W_i} Kb/s, and

17     $\cos t_{X_{W_i}}^{disk}$ is a cost of consumed resources for a stream having a disk access to a file encoded at

18   X_{W_i} Kb/s .

1   11.   The method of claim 1 further comprising:

2   receiving at least one service parameter.

1   12.   The method of claim 11 wherein said at least one service parameter comprises

2 information identifying at least one performance criteria desired to be satisfied by said at least

3 one streaming media server under the workload.

1   13.   The method of claim 12 wherein said at least one performance criteria specifies a

2 minimum percentage of time that said at least one streaming media server is desired to be

3 capable of supporting the workload.

1   14.   The method of claim 11 wherein said at least one service parameter comprises

2 information identifying a constraint.

1   15.   The method of claim 11 wherein said evaluating further comprises:

2   evaluating whether said at least one streaming media server satisfies said at least

3 one service parameter.

1   16.   Computer-executable software code stored to a computer-readable medium, the

2 computer-executable software code comprising:

3   code for receiving workload information for a workload of client accesses of

4 streaming media files from a server; and

5   code for employing a cost function derived for at least one system configuration

6 from a single file benchmark and a unique file benchmark for evaluating a capacity of the at least

7 one system configuration for supporting the workload.

1   17.   Computer-executable software code of claim 16 further comprising:

2   code for receiving configuration information for said at least one system

3 configuration.

1   18.   Computer-executable software code of claim 16 wherein said code for evaluating

2   a capacity of at least one system configuration for supporting the workload comprises:

3   code for determining whether said at least one system configuration is capable of

4   supporting said workload in accordance with at least one service parameter.

1   19.   Computer-executable software code of claim 18 wherein said at least one service

2   parameter comprises information identifying at least one performance criteria desired to be

3   satisfied by said at least one system configuration under the workload.

1   20.   Computer-executable software code of claim 16 further comprising:

2   code for generating a workload profile for the received workload information.

1   21.   Computer-executable software code of claim 20 wherein the received workload

2   information comprises an access log collected over a period of time.

1   22.   Computer-executable software code of claim 20 wherein said workload profile

2   comprises:

3   for a plurality of different points in time, identification of a number of concurrent

4   client accesses, wherein the number of concurrent client accesses are categorized into

5   corresponding encoding bit rates of streaming media files accessed thereby and are further sub-

6   categorized into either memory or disk accesses.

1   23.   Computer-executable software code of claim 16 wherein said code for

2   evaluating comprises:

3   code for generating a service demand profile for said at least one system

4   configuration.

1   24.   Computer-executable software code of claim 16 wherein said code for evaluating

2   a capacity of at least one system configuration comprises:

3   code for evaluating a capacity of a plurality of different system configurations and

4   determining an optimal one of said plurality of different system configurations for supporting

5   the workload.

1   25.   A system comprising:

2   means for receiving configuration information for a plurality of different system

3   configurations, wherein the configuration information comprises, for each of the plurality of

4   different system configurations, a corresponding single file benchmark and unique file

5   benchmark, wherein said single file benchmark measures capacity of the corresponding system

6   configuration for serving a population of clients that all access a same file, wherein said unique

7   file benchmark measures capacity of the corresponding system configuration for serving a

8   population of clients that all access different files;

9   means for receiving workload information for a workload of client accesses of

10   streaming media files from a server; and

11   means for evaluating, based on the configuration information, the capacity of each

12   of said plurality of different system configurations for supporting said workload.

1   26.   The system of claim 25 further comprising:

2   means for determining an optimal one of said plurality of different system

3   configurations for supporting said workload.

1   27.   The system of claim 26 wherein said means for determining an optimal one of

2   said plurality of different system configurations for supporting said workload determines a most

3   cost-effective one of said plurality of different system configurations for supporting said

4   workload according to determined service parameters.

1   28.   A method comprising:

2   receiving workload information identifying an expected workload of client

3 accesses of streaming media files from a server over a period of time; and

4   determining a service demand profile for at least one server configuration under

5 evaluation for evaluating a capacity of said at least one server configuration for supporting the

6 expected workload, wherein said service demand profile comprises a plurality of pairs of

7 information, each pair comprising an identification of a duration of time in said period of time

8 and a corresponding computed resource cost of the at least one server configuration for serving

9 the workload over the duration of time.

1   29.   The method of claim 28 further comprising:

2   receiving at least one service parameter.

1   30.   The method of claim 29 wherein said at least one service parameter comprises

2 information identifying at least one performance criteria desired to be satisfied by said at least

3 one server configuration under the expected workload.

1   31.   The method of claim 29 further comprising:

2   evaluating the determined service demand profile for the at least one server

3 configuration to determine whether the at least one server configuration satisfies the received at

4 least one service parameter.

1   32.   A system comprising:

2   a media profiler operable to receive a client access log collected over a period of

3 time for a service provider's site and generate a workload profile for the service provider's site,

4 wherein said workload profile comprises, for a plurality of different points in time, identification

5 of a number of concurrent client accesses, wherein the number of concurrent client accesses are

6 categorized into corresponding encoding bit rates of streaming media files accessed thereby and

7 are further sub-categorized into either memory or disk accesses; and

8   a capacity evaluator operable to receive the generated workload profile and

9 evaluate at least one server configuration's capacity for supporting the site's workload.

1   33   The system of claim 32 wherein said capacity evaluator is further operable to

2   receive configuration information for said at least one server configuration.

1   34.   The system of claim 32 wherein in evaluating said at least one server

2   configuration's capacity, said capacity evaluator determines whether said at least one server

3   configuration is capable of supporting the site's workload in accordance with at least one service

4   parameter.

1   35.   The system of claim 34 wherein said at least one service parameter comprises

2   information identifying at least one performance criteria desired to be satisfied by said at least

3   one server configuration under the site's workload.

1   37.   The system of claim 32 wherein in evaluating said at least one server

2   configuration's capacity, said capacity evaluator is operable to generate a service demand profile

3   for said at least one server configuration.

1   38.   The method of claim 1 further comprising:

2   deriving, by said capacity planning tool, from the single file benchmark and

3   unique file benchmark, a cost function for measuring the capacity of the at least one streaming

4   media server for supporting the workload.

1   39.   The method of claim 1 wherein said evaluating comprises computing a service

2   demand for said at least one streaming media server supporting said workload; and

3   wherein said computing said service demand comprises computing:

4

$$Demand = \sum_{i=1}^{K_W} N_{X_{W_i}}^{memory} \times \cos t_{X_{W_i}}^{memory} + \sum_{i=1}^{K_W} N_{X_{W_i}}^{disk} \times \cos t_{X_{W_i}}^{disk} \; ,$$

5   wherein the workload W comprises $X_W = X_1,...,X_{k_W}$ set of different encoded bit

6   rates of files served in the workload, $N_{X_{W_i}}^{memory}$ is a number of streams in the workload having a

7   memory access to a subset of files encoded at X_{W_i} Kb/s, $\cos t_{X_{W_i}}^{memory}$ is a cost of consumed

8   resources for a stream having a memory access to a file encoded at X_{W_i} Kb/s, $N_{X_{W_i}}^{disk}$ is a number

9   of streams in the workload having a disk access to a subset of files encoded at X_{W_i} Kb/s, and

10    $\cos t_{X_{W_i}}^{disk}$ is a cost of consumed resources for a stream having a disk access to a file encoded at

11   X_{W_i} Kb/s .

1   42.   The method of claim 28 further comprising:

2   deriving, from a single file benchmark and unique file benchmark of the at least

3   one server configuration, a cost function for computing resource cost of the at least one server

4   configuration; and

5   employing said cost function for computing the computed resource cost of the at

6   least one server configuration for serving the workload over the duration of time.

1   43.   The system of claim 37 wherein said service demand profile comprises a plurality

2   of pairs of information, each pair comprising identification of a duration of time in said period of

3   time and a corresponding computed resource cost of the at least one server configuration for

4   serving the workload over the duration of time.

1   44.   A method comprising:

2   determining results of a single file benchmark for each of a plurality of encoding

3 bit rates of a single file served by at least a first streaming media server configuration, wherein

4 the result of the single file benchmark for a given encoding bit rate identifies the maximum

5 number of concurrent streams of the single file that the at least a first streaming media server

6 configuration can supply to a population of clients at the given encoding bit rate;

7   determining results of a unique file benchmark for each of said plurality of

8 encoding bit rates, wherein the result of the unique file benchmark for a given encoding bit rate

9 identifies the maximum number of concurrent streams of different files that the at least a first

10 streaming media server configuration can supply to the population of clients at the given

11 encoding bit rate;

12   deriving, from the results of the single file benchmark and unique file benchmark,

13 a cost function;

14   receiving, into a capacity planning tool, workload information for a workload of

15 client accesses of streaming media files from a server; and

16   using, by the capacity planning tool, the cost function for said at least a first

17 streaming media server configuration for evaluating a capacity of the at least a first streaming

18 media server configuration for supporting the workload.

IX.   <u>EVIDENCE APPENDIX</u>

None.

X.   <u>RELATED PROCEEDINGS APPENDIX</u>

None.